Precision of Prediction in Second Order Calibration, with Focus on Bilinear Regression Methods.

Marie Linder, Rolf Sundberg*

Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

October 24, 2000

SUMMARY

We consider calibration of hyphenated instruments with particular focus on determination of the unknown concentrations of new specimens. A hyphenated instrument generates for each specimen a two-way array of data. These are assumed to depend on the concentrations through a bilinear regression model, where each constituent is characterized by a pair of profiles to be determined in the calibration. We discuss the problem of predicting the unknown concentrations in a new specimen, after calibration. We formulate three different predictor construction methods, a "naïve" method, a least squares method, and a refined version of the latter that takes account of the calibration uncertainty. We give formulae for the uncertainty of the predictors under white noise, when calibration can be seen as precise. We refine these formulae to allow for calibration uncertainty, in particular when calibration is carried out by the bilinear least squares (BLLS) method or the singular value decomposition (SVD) method proposed by Linder and Sundberg (*Chemometrics Intell. Lab. Syst.*, **42**, 159–178 (1998)). By error propagation formulae and previous results on the precision of $\widehat{A}$ and $\widehat{B}$ we can obtain approximate standard errors for the predicted concentrations, according to each of the two estimation methods. The performance of the predictors and the precision formulae is illustrated on both real (fluorescence) and simulated data.

KEY WORDS: bilinear regression, BLLS estimation method, calibration, fluorescence data, hyphenated instruments, LS prediction, PARAFAC, prediction, second order calibration, SVD estimation method

1 Introduction

For efficient quantitative analysis of multicomponent systems in analytical chemistry, hyphenated instruments can be used, generating two-dimensional arrays of data for each specimen. For many such instruments, including LC–UV, LC–MS, GC–MS, GC–FTIR and excitation-emission fluorescence, this array has a bilinear structure. This bilinear structure motivates a bilinear regression model for the dependence on concentrations. It is not evident how determination/prediction of new specimens should best be carried out through a calibrated bilinear regression. We will discuss

*Correspondence to: Rolf Sundberg, Mathem. Statistics, Stockholm University, SE-106 91 Stockholm, Sweden
E-mail rolfs@matematik.su.se
Contract/grant sponsor: Swedish Natural Science Research Council

different prediction methods suited for this situation and give formulae for their uncertainty, and we will combine these prediction methods with two different calibration methods. A theoretical statistical analysis will be supplemented by a simulation study and an illustration on excitation–emission data.

1.1 Second order calibration

Sanchez and Kowalski [1] coined the terms zero, first and second order calibration. Zero order calibration is the univariate situation, represented by for example a monochromatic photometer and a single constituent to be quantified. The classical statistical calibration procedure is to fit a linear regression of instrument response (absorbance, say) on concentration, motivated by the Lambert–Beer law, and invert this relationship to determine the concentration in a new specimen. However, in "natural calibration", when all specimens are regarded as sampled from some natural population, the classical procedure should be replaced by a truly predictive point of view. The best predictor is given by the direct regression of concentration on absorbance, i.e. regression the other way round. First order calibration is the multi-dimensional extension to several wavelengths and one or several constituents. The predictive point of view is here dominating in practice, that is concentrations are regressed on the spectral data, if not directly as in OLS and ridge regression, so via some sort of latent variable representation as in PLS and PCR. However, there are alternative multivariate regression methods for regression of the spectral data on the concentrations, extending the classical approach from the univariate case. Such methods might be motivated if the calibration specimens are strongly controlled. The richer data from first order calibration allow interferents to be detected during prediction, but not to be corrected for. The data arrays of hyphenated instruments are even richer, and one calibration specimen is in principle enough to determine several constituents in a new specimen (GRAM [2]). Also, in second order calibration it is possible not only to detect, but also to correct for the presence of interfering species [1]. This so called second order advantage is occasionally even found as part of the definition of second order calibration.

In second order calibration three different principles can be used for determination (prediction) of a new concentration. If the classical approach is followed, the bilinear regression should first be calibrated on some specimens of known composition, and the fitted bilinear regression model inverted for the determination of the new concentrations. Linder & Sundberg [3] studied two methods for the calibration phase, *bilinear least squares* (BLLS) and what they called the *singular value decomposition* (SVD) methods of estimation. This principle is followed and investigated in the present paper. Different inversion methods will be discussed, and the most naïve one will be found theoretically inefficient under white noise.

Another two-stage principle is represented by methods of PARAFAC type, see Bro [4] for a review (in particular its Section 10). The idea is then first to fit a model with all concentrations regarded as unknowns, and next for the prediction phase to calibrate the PARAFAC solution against known concentrations by multiple regression of scores ("estimated" concentrations) on true concentrations. Two designs of PARAFAC are obtained depending on whether the unknown specimens are included already in the model fit, or predicted separately. The present approach is relevant for the latter design. This design is necessary if specimens arrive sequentially after the calibration, but could be advocated also in other cases.

The third principle is represented by PLS and PCR type procedures, where the concentrations are regressed on the instrument data [5,6,7]. In their simplest versions these methods do not utilize the bilinear structure of the instrument data. This implies that they do not yield estimates of the true profiles (curve resolution) for the two modes of the instrument. They also (implicitly at

least) assume that the calibration specimens come from the same natural population as the new specimen. This is for example not satisfied in the illustration of Section 7 below.

All these methods and method types have their pros and cons. Their statistical efficiencies are compared in a large scale simulation study in a separate paper by Linder [8]. Anticipating the results found there, the bilinear least squares (BLLS) regression method for calibration in combination with one of the inversion methods to be designed in the present paper, usually will have the highest statistical efficiency of all methods. It might come as a surprise that PARAFAC is generally less efficient, since PARAFAC is also based on least squares, albeit in two steps, and standard PARAFAC uses all data already in the model fitting. One explaining factor is that PARAFAC is less parsimonious than bilinear regression, because of the multiple regressions in the calibration of it. However, with calibration by simple proportional regression instead of multiple, PARAFAC is parsimonious but much less efficient. The SVD method [3], as an alternative, is somewhat less efficient than BLLS, but is theoretically simpler and easier to use. Another important advantage of both these methods to their competitors is the availability of precision formulae in the form of estimated standard errors and variance–covariance matrices for the instrument profiles. These results will be used in the present paper to develop precision formulae for the predicted composition of a new specimen, taking account of the calibration uncertainty. With other calibration/prediction methods like PARAFAC, TLD [9] and PLS-type methods, precision formulae are lacking. GRAM [2] is an exception, for which Faber et al [10] derived standard errors. Indirect quantification of type cross-validation, which is the standard tool in first order, is not a good idea when the calibration involves only few specimens. On the other hand, the precision formulae of course require that the statistical model is reasonably adequate. An algorithm will be proposed that may correct for one type of model error, namely an interferent in the new specimen (i.e. the second order advantage).

The number of calibration specimens is typically small. At least for the calibration part of the problem we prefer the word specimen, rather than sample, since when calibration specimens are few they will typically have been generated systematically, rather than by random sampling. For the new specimen of unknown concentrations we will use the term prediction rather than estimation for the determination of the composition of the new sample, even if the composition of the new specimen is not regarded as inherently random.

Next we describe the bilinear regression model which the data from the hyphenated instrument are assumed to follow. In Section 2 we then suppose that calibration using chemical mixtures of known composition has been carried out separately, and we introduce and compare three different methods of prediction. General formulae for their approximate precision are given in Section 3. A procedure to correct for interferents is suggested in Section 4. In Section 5 we give a brief account of two methods (BLLS and SVD) for the calibration part of the problem. In Sections 6 and 7 comparisons of methods are carried out on simulated and on real data, and the precision formulae are tested against data.

1.2 The bilinear regression model

To have a hyphenated instrument in mind, we use LC–UV as model example. Ideally its response should follow a bilinear model, albeit in practice there can be chromatographic synchronization problems and other causes of bilinearity deviations. Let us consider how data are generated in an LC–UV instrument. First the constituents are partially separated in time by the LC, and then a UV spectrum is recorded for each chromatographic time, so we get a matrix or array of absorbances, a spectro-chromatogram, indexed by time and wavelength for each specimen. Each constituent is characterized by its elution profile and its UV-spectrum profile.

For each pure constituent, the Lambert–Beer law implies that the absorbance spectrum is proportional to the concentration. Theory also says that elution peak areas are proportional to concentration. We thus assume that concentrations are not so high that they violate these laws, and that the response is additive over constituents. Also we assume there are no chromatographic synchronization problems, and that the background has been eliminated; else we would have to add some type of intercept parameter. We will denote the absorbance z at time i and wavelength j in specimen k by $z_{ij}^{(k)}$, forming matrices $Z^{(k)}$; constituent will be indexed by $r = 1, \ldots, R$.

The proportionality in each instrument response implies that a pure constituent should yield a rank 1 response matrix of type $c_{rr}\alpha_r\beta_r^T$, where c_{rr} is the concentration and the column vectors α_r and β_r represent the chromatographic and spectrum profiles for constituent r. The additivity over the R constituents implies that a mixture of them will yield a rank R response matrix Z, that can be expressed as $Z^{(k)} = \sum_{r=1}^{R} c_{rr}^{(k)}\alpha_r\beta_r^T$ with elements $z_{ij}^{(k)} = \sum_{r=1}^{R} c_{rr}^{(k)}\alpha_{ir}\beta_{jr}$ for specimen k. In matrix form this model can be written

$$\textbf{Model for data: } Z^{(k)} = AC^{(k)}B^T + E^{(k)}. \tag{1}$$

Here A and B have columns α_r and β_r, respectively, r=1,...,R, $C^{(k)}$ is an $R \times R$ diagonal matrix with the concentrations $c_{rr}^{(k)}$ along the diagonal, and $E^{(k)}$ represents noise, k=1,...,K. When it comes to precision results below, we assume additive white noise with variance σ^2. For examples of data matrices simulated under this model, see Figure 1.

An alternative form of the model would be obtained by writing the whole set of K two-way arrays as a three-way array, with three exchangeable "ways" (modes), as in PARAFAC. However, this would obscure the special role of the concentrations in a calibration–prediction situation, so we will keep to the form (1).

Model (1) is not identifiable, since we can choose an arbitrary scaling for α_r and β_r. However, as long as we do not need indentifiability, we will use this model representation (1). When it comes to precision of estimation of the model parameters, we will need unique parameters. To this end we introduce scale parameters γ_{rr} and the constraints $|\alpha_r| = |\beta_r| = 1$ to get the constrained model representation $Z^{(k)} = \sum_{r=1}^{R} c_{rr}^{(k)}\gamma_{rr}\alpha_r\beta_r^T$. In matrix form this modified formulation is

$$\textbf{Constrained model representation: } Z^{(k)} = AC^{(k)}\Gamma B^T + E^{(k)},\ |\alpha_r| = |\beta_r| = 1. \tag{2}$$

The identifiability of this model follows from the constraint that $C^{(k)}$ and Γ are both diagonal matrices, and that no rotation of A or B can be done without destroying this property.

2 Prediction methods

We here discuss the prediction phase from a relatively general point of view, without specifying the calibration method used, but assuming that calibration has been carried out, resulting in estimated profiles $\widehat{A}$ and $\widehat{B}$ of model (1). A new specimen of unknown diagonal concentration matrix C_0, with concentration vector c_0 along the diagonal, yields a response matrix Z_0, which is supposed to follow the same bilinear model (1) as the calibration samples, but with unknown c_0:

$$Z_0 = AC_0B^T + E_0. \tag{3}$$

We will formulate three different methods for determination of the new composition c_0. This is because these methods all have their advantages and disadvantages, and because they can be regarded as forming a sequence of successive refinements. We will discuss their relationships, and in later sections give formulae for their precision and illustrate the use of them.

2.1 The naïve prediction method

An intuitively natural way to form a predictor of the diagonal c_0 of the concentration matrix C_0 is to start by largely eliminating A and B from the right hand side of (3), by multiplication from left and right by suitable inverses of the corresponding estimated coefficient matrices. We use left inverses $\widehat{A}^-$ and $\widehat{B}^-$ of $\widehat{A}$ and $\widehat{B}$, respectively, such that (by definition) $\widehat{A}^-\widehat{A} = \widehat{B}^-\widehat{B} = I$. This results in an approximately diagonal matrix of form,

$$\widetilde{C}_0 = \widehat{A}^- Z_0 \widehat{B}^{-T} = \widehat{A}^- A C_0 B^T \widehat{B}^{-T} + \widehat{A}^- E_0 \widehat{B}^{-T}. \tag{4}$$

If calibration has been precise, E_0 in the last term is the dominating source of randomness, and then we aim at selecting left inverses that minimize the variances of the elements of this term, under the assumption of uncorrelated homoscedastic noise (for simplicity). In Appendix A we show the not very surprising result that the best choices are the Moore–Penrose generalized inverses, $\widehat{A}^+$ and $\widehat{B}^+$. In the sequel, by the naïve predictor $\tilde{c}_0$ we will mean the diagonal of $\widetilde{C}_0$ in (4), furnished with the Moore–Penrose inverses, that is

$$\tilde{c}_0 = Diag[\widetilde{C}_0] = Diag[\widehat{A}^+ Z_0 \widehat{B}^{+T}]. \tag{5}$$

2.2 The least squares prediction method

The concentration matrix C_0 is diagonal by definition, but the matrix $\widetilde{C}_0$ computed from data as in (4) will not be exactly diagonal. The off-diagonal elements can actually contain much useful information about the errors in $\hat{c}_{rr}$, because the elements of $\widetilde{C}_0$ are more or less strongly correlated. One method that utilizes the whole of matrix (4) is the least squares method, neglecting the calibration errors in the estimated profiles $\widehat{A}$ and $\widehat{B}$. With unknown parameters only in $c_0 = \mathrm{Diag}[C_0]$ we have a linear model for Z_0 in terms of these R concentrations. The normal equations for OLS estimation in this linear model are obtained by identifying the diagonal elements of $\widehat{A}^T Z_0 \widehat{B}$ with their expected values,

$$\mathrm{Diag}[\widehat{A}^T Z_0 \widehat{B}] = \mathrm{Diag}[(\widehat{A}^T \widehat{A}) C_0 (\widehat{B}^T \widehat{B})]. \tag{6}$$

Solving this equation system for c_0 yields the desired LS predictor. With the notation $\odot$ for the elementwise (or Schur, or Hadamard) product of two matrices, the solution of (6) can be explicitly written down as

$$\hat{c}_0 = \{(\widehat{A}^T \widehat{A}) \odot (\widehat{B}^T \widehat{B})\}^{-1} \mathrm{Diag}[\widehat{A}^T Z_0 \widehat{B}], \tag{7}$$

Since the elementwise product is directly available in for example MATLAB, the form (7) is easily programmed for the computer.

In order to see how the LS predictor (7) is related to the naïve predictor $\tilde{c}_0$, note that the right-most factor of (7) may also be written

$$\mathrm{Diag}[\widehat{A}^T Z_0 \widehat{B}] = \mathrm{Diag}[(\widehat{A}^T \widehat{A}) \widetilde{C}_0 (\widehat{B}^T \widehat{B})]. \tag{8}$$

The right hand side here shows how the LS predictor makes use of the whole matrix $\widetilde{C}_0$, and not only of its diagonal.

2.3 A refined least squares prediction method

The simple LS predictor above is best (minimum variance) linear unbiased (BLUE) under the assumption of no uncertainty in the profiles (and the z_{ij} having homoscedastic white noise). If we know (or have formulae for) the profile uncertainties, a refinement c_0^* can be constructed theoretically from a regression of each diagonal element in $\widetilde{C}_0$ on the set of off-diagonal elements. This regression will then utilize the covariance matrix for $\widetilde{C}_0$, to be found in the next section.

To demonstrate how the regression is carried out, in principle, we provisionally introduce the notation X for the set of off-diagonal elements of $\widetilde{C}_0$, that is $X = \{\tilde{c}_{r_1 r_2}\}_{r_1 \neq r_2}$. Denoting the variance–covariance matrix for $(\tilde{c}_{rr}, X)$ by

$$\begin{pmatrix} \Sigma_{rr} & \Sigma_{rx} \\ \Sigma_{xr} & \Sigma_{xx} \end{pmatrix},$$

see Result 3.3 for explicit formulae, the linear regression of $\tilde{c}_{rr}$ on X is the expected value

$$\mathrm{E}[\tilde{c}_{rr}|X] = c_{rr} + \Sigma_{rx}\Sigma_{xx}^{-1}X.$$

This motivates the refinement

$$c_{rr}^* = \tilde{c}_{rr} - \Sigma_{rx}\Sigma_{xx}^{-1}X. \tag{9}$$

The precision of the predictor will be reduced from $\mathrm{V}(\tilde{c}_{rr}) = \Sigma_{rr}$ to

$$\mathrm{V}[c_{rr}^*|X] = \Sigma_{rr} - \Sigma_{rx}\Sigma_{xx}^{-1}\Sigma_{xr} \tag{10}$$

Primarily this refinement can be worth-while if there is considerable calibration uncertainty. If calibration uncertainty is negligible we just get the LS predictor again. This is most easily seen indirectly, from the fact that both are based on the same data and have minimum variance.

3 Precision formulae for the predictors

The precision of the naïve predictor $\tilde{c}_0$ and the LS predictor $\hat{c}_0$ are expressed by the variance–covariance matrices of the stochastic vectors (5) and (7), respectively. For the refined LS predictor c_0^* of the previous section, already the construction requires the variance–covariance matrix for $\widetilde{C}_0$. We start by giving the relatively simple precision formulae when calibration errors can be neglected. In this case $\hat{c}_0$ and c_0^* are the same. From now on all results will be expressed in terms of the constrained model representation (2), that is with B replaced by $B\Gamma$, and the columns of B having unit length. The parameterization uniqueness implied by this constraint will be necessary from Result 3.2 onwards. Except in the most general Result 3.3, we also assume from now on that measurement noise is white with variance σ^2. As before $\odot$ denotes the elementwise matrix product.

Result 3.1: Prediction variance–covariance matrices when calibration is precise
When calibration errors can be neglected, the variance–covariance matrices for the naïve and LS predictors are given by the formulae

$$\begin{aligned} V[\tilde{c}_0] &= \sigma^2 \, (A^T A)^{-1} \odot (\Gamma B^T B \Gamma)^{-1}, \\ V[\hat{c}_0] = V[c_0^*] &= \sigma^2 \, \{(A^T A) \odot (\Gamma B^T B \Gamma)\}^{-1}. \end{aligned}$$

Justification: For the naïve predictor $\tilde{c}_0$, this result is essentially included in the proof of the

optimality of the Moore–Penrose inverse, given in Appendix A, see (14) and (15). More generally, the result is just an application of the rules for calculating variances and covariances of linear expressions, in contrast with the nonlinear results to follow. Details are omitted. ■

The following results will incorporate the statistical uncertainty from the calibration. We first give the relatively simple variances of the naïve predictor (5), as yet not with any specific calibration method in mind.

Result 3.2: General prediction variances, naïve predictor
The prediction variance for the concentration of constituent r is

$$\begin{aligned} V[\tilde{c}_{rr}] &\approx c_{rr}^2\{(A^+ V[\hat{\alpha}_r]A^{+T})_{rr} + (B^+ V[\hat{\beta}_r]B^{+T})_{rr} + (\Gamma^{-1} V[\hat{\gamma}_{rr}]\Gamma^{-1})_{rr} \\ &+ 2[(A^+ Cov[\hat{\alpha}_r, \hat{\beta}_r]B^{+T})_{rr} + (\Gamma^{-1}A^+ Cov[\hat{\alpha}_r, \hat{\gamma}_{rr}])_r + (\Gamma^{-1}B^+ Cov[\hat{\beta}_r, \hat{\gamma}_{rr}])_r]\} \\ &+ \sigma^2 \{(A^T A)^{-1}\}_{rr}\{(\Gamma B^T B\Gamma)^{-1}\}_{rr}. \end{aligned}$$

Justification: For a mathematical justification, see Appendix B. ■
Remark: Note that all terms due to the calibration uncertainty are proportional to the concentration squared, c_{rr}^2. ■

We now turn to the whole variance–covariance matrix for the matrix $\widetilde{C}_0$. To be able to write down the variance–covariance matrix for $\widetilde{C}_0$, we vectorize this matrix. We use the conventions $\mathrm{V}[M] = \mathrm{E}[\mathrm{Vec}(M)\mathrm{Vec}(M)^T] - \mathrm{E}[M]\mathrm{E}[M]^T$ for the variance of a matrix M, and analogously $\mathrm{Cov}[M, N] = \mathrm{E}[\mathrm{Vec}(M)\mathrm{Vec}(N)^T] - \mathrm{E}[M]\mathrm{E}[N]^T$ for the covariance between matrices M and N, where $\mathrm{E}[M] = \mathrm{E}[\mathrm{Vec}(M)]$. The following rules can be established for the calculation of variances and covariances for compound matrices as follows, see Graham [11]:

$$\begin{aligned} \mathrm{V}[AMB] &= (B^T \otimes A)\mathrm{V}[M](B^T \otimes A)^T \\ \mathrm{Cov}[AMB, CND] &= (B^T \otimes A)\mathrm{Cov}[M, N](D^T \otimes C)^T \\ \mathrm{V}[M^T] &= U\mathrm{V}[M]U^T \\ \mathrm{Cov}[M^T, N] &= U\mathrm{Cov}[M, N] \\ \mathrm{Cov}[M, N^T] &= \mathrm{Cov}[M, N]U^T, \end{aligned}$$

where $\otimes$ is the direct (or Kronecker) matrix product and U is the permutation matrix associating $\mathrm{Vec}(M)$ and $\mathrm{Vec}(M^T)$, i.e. $\mathrm{Vec}(M^T) = U\mathrm{Vec}(M)$ defines U.

Now we can present the results from application of these rules to yield the variance–covariance matrix of $\widetilde{C}_0$. It is obtained by analogous calculations to those of the special case above, Result 3.2.

Result 3.3: General variance–covariance matrix for the naïve predictor

$$\begin{aligned} V[\widetilde{C}_0] \approx\ & (C_0 \otimes A^+)\, V[\widehat{A}](C_0 \otimes A^+)^T + \\ & (C_0 \otimes A^+)\, Cov[\widehat{A}, \widehat{B}]U^T(\Gamma^{-1}B^+ \otimes C_0\Gamma)^T + \\ & (C_0 \otimes A^+)\, Cov[\widehat{A}, \widehat{\Gamma}](\Gamma^{-1} \otimes C_0)^T + \\ & (\Gamma^{-1}B^+ \otimes C_0\Gamma)U\, Cov[\widehat{B}, \widehat{A}](C_0 \otimes A^+)^T + \\ & (\Gamma^{-1}B^+ \otimes C_0\Gamma)U\, V[\widehat{B}]U^T(\Gamma^{-1}B^+ \otimes C_0\Gamma)^T + \\ & (\Gamma^{-1}B^+ \otimes C_0\Gamma)U\, Cov[\widehat{B}, \widehat{\Gamma}](\Gamma^{-1} \otimes C_0)^T + \\ & (\Gamma^{-1} \otimes C_0)\, Cov[\widehat{\Gamma}, \widehat{A}](C_0 \otimes A^+)^T + \end{aligned}$$

$$(\Gamma^{-1} \otimes C_0)\, Cov[\widehat{\Gamma}, \widehat{B}] U^T (\Gamma^{-1} B^+ \otimes C_0 \Gamma)^T +$$
$$(\Gamma^{-1} \otimes C_0)\, V[\widehat{\Gamma}] (\Gamma^{-1} \otimes C_0)^T +$$
$$(\Gamma^{-1} B^+ \otimes A^+)\, V[E_0] (\Gamma^{-1} B^+ \otimes A^+)^T. \qquad (11)$$

This is a long expression, but it is quite convenient for computer implementation. Note also that all terms except the last one represent the calibration uncertainty.

The least squares and refined least squares predictors $\hat{c}_0$ and c_0^* are linear in the elements of $\widetilde{C}_0$, see (7) and (8) for $\hat{c}_0$ and (9) for c_0^*. Admittedly the coefficients in these linear expressions involve the estimated parameters from the calibration, but if the uncertainty in these coefficients is neglected the variance–covariance matrices of the predictors are linearly given in terms of the elements of $V(\widetilde{C}_0)$. For c_0^*, see (10) for the explicit expression which was used in the simulation study to be described in Section 6.

4 The second order advantage.

As mentioned in the Introduction, second order calibration has the advantage that the presence of a new substance, that was not calibrated for, in a specimen whose contents are to be determined, can be detected and corrected for. We have designed an algorithm that may add such a second order advantage to our proposed prediction methods. As also in for example PARAFAC, interfering constituents can be found by fitting more components than calibrated for.

Our algorithm proceeds as follows, after a rank check on Z_0 or a smallness check on the nondiagonal elements of $\widetilde{C}_0$ has indicated the presence of an interfering constituent in the new specimen. Given estimates of the profiles and the scaling matrix, $\widehat{A}_0$ and $\widehat{B}_0$ and $\widehat{\Gamma}_0$, and the new specimen data Z_0, predict its diagonal concentration matrix C_0 by the selected predictor, here denoted $\widehat{C}_0$. Then repeatedly go through the following steps until convergence or divergence:

1. Calculate the matrix of nonnegative residuals, $R_+ = \max\{0, Z_0 - \widehat{A}_0 \widehat{C}_0 (\widehat{B}_0 \widehat{\Gamma}_0)^T\}$
2. Make a singular value decomposition of R_+ to find its first singular vectors u_0 and v_0, corresponding to the largest singular value. These are estimates of the profiles of the interfering constituent.
3. Extend the original estimated profiles $\widehat{A}_0$ and $\widehat{B}_0$ with the profiles u_0 and v_0, and predict a correspondingly extended c_0 from $(\widehat{A}_0, u_0)$ and $(\widehat{B}_0 \widehat{\Gamma}_0, v_0)$
4. Restrict the extended c_0 to the original size, form the corresponding $\widehat{C}_0$, and go back to the first step to calculate a new R_+.

If there is a substantial contribution from an extra substance, that clearly exceeds the noise level, the algorithm can be expected to pick out this substance. Our limited experience tells that the algorithm works well when there is relatively little noise in the data. The different prediction methods are not equally sensitive to an interfering substance, and also the algorithm will converge at different speed for different methods, but the results are too premature to be reported in any further detail in the present paper.

5 Two profiles estimation procedures

In Section 3 formulae were given for the precision of various predictors of the concentrations in a new specimen, not only when calibration was assumed precise, but also when attention was

paid to the calibration uncertainty, in general terms. Here we present two different calibration methods which allow us to specify this calibration uncertainty. These methods are the bilinear least squares (BLLS) method and what we have called the singular value decomposition (SVD) method, both introduced and described in detail in Linder and Sundberg [3]. We here just give a sketchy introduction to these methods and refer to Appendix C for some more details, and in particular for the precision formulae we require.

In bilinear least squares estimation we fit the bilinear regression model by minimizing the sum of squared deviations between all the elements of all the data matrices $Z^{(k)}$ and their expected values according to the model (1) or (2). Differentiation yields the normal equations. It turns out that these involve data only in terms of the R matrix statistics

$$T^{(r)} = \sum_{k=1}^{K} c_{rr}^{(k)} Z^{(k)}, \tag{12}$$

one for each constituent. These are concentration-weighted sums of the primary $Z^{(k)}$ arrays. The normal equations are obtained by forming the parameter-weighted sums of the rows and columns of these matrices, $T^{(r)}\beta_r$ and $\alpha_r^T T^{(r)}$, and set them to their expected values, see Result C.1. Because the weights are parameter-dependent in this way, the normal equations cannot be solved explicitly, but in analogy with PARAFAC an alternating least squares (ALS) algorithm can be used, and it usually works without problems. The analogy with PARAFAC can be made explicit by saying that BLLS on the calibration data is the same as PARAFAC with one mode fixed, cf. Kroonenberg [12], chapter 5, pp 113-114.

The SVD method refrains from forming the parameter-weighted sums by instead setting the $T^{(r)}$s themselves to their expected values. The resulting equation system can be reexpressed and shown to be equivalent with R singular value decompositions, each estimating the corresponding pure component rank 1 matrix $\alpha_r\beta_r^T$ (or $\gamma_{rr}\alpha_r\beta_r^T$ under the constraints $|\alpha_r| = |\beta_r| = 1$), see Result C.3.

As for precision, the variance–covariance matrix of the BLLS estimates cannot be exactly given, since the model is nonlinear, but an approximate version can be obtained by local linearization of model (2) and of its constraints, see Result C.2. For the SVD method an explicit approximate variance–covariance matrix can be written down, see Result C.4. These results are derived under the assumption of white noise with variance σ^2. For both methods the approximations are good if σ is small enough or the number of adequate calibration specimens is large enough.

The BLLS and SVD methods assume that the calibration is done with at least as many specimens as constituents. This condition is likely to be satisfied in most cases, but it is not really necessary for the quantitative prediction of the same constituents in a new specimen, as shown by the GRAM method [2]. It is also not always necessary for identifiability in the estimation problem [13].

In principle the least squares estimation method could refer to all data jointly, if only the unknown concentrations of new specimens are regarded as additional parameters to be estimated, besides the profiles. That is, what we here have referred to as separate estimation and prediction phases would then be joined in one single optimization procedure, more like standard PARAFAC. We have not gone this way. Both numerically and theoretically the complications would have been considerably increased, but the properties not necessarily improved, cf. the negative results by Brown and Sundberg [14] on the analogous question in first order calibration, and simulation results on PARAFAC by Linder [8] not indicating any improvement with the joint procedure.

Use of the explicit expressions for the variances and covariances of the SVD estimator yields the following relatively simple formula for the approximate precision of the predicted concentration

$\tilde{c}_{rr}$ by the naïve method. Here $(D^{-1})_{rr}$ is the rth diagonal element of the inverse of the $R \times R$ sums of concentration products matrix D, which has elements $\sum_k c^{(k)}_{r_1 r_1} c^{(k)}_{r_2 r_2}$.

Result 5.1: Prediction variance for the naïve predictor with the SVD estimator
The variance of the naive predictor $\tilde{c}_{rr}$ for constituent r, considering the uncertainty of the SVD estimation method, is given approximately by

$$V[\tilde{c}_{rr}] \approx \frac{\sigma^2 c_{rr}^2 (D^{-1})_{rr}}{\gamma_{rr}^2} [\{(A^T A)^{-1}\}_{rr} + \{(B^T B)^{-1}\}_{rr} - 1 + \frac{\{(A^T A)^{-1}\}_{rr} \{(B^T B)^{-1}\}_{rr}}{c_{rr}^2 (D^{-1})_{rr}}]. \quad (13)$$

A justification of this formula is given towards the end of Appendix C. The result is used both in the simulation study, where the influence of different factors is discussed, and in the real data illustration below. It can also be regarded as a relatively explicit example of how different factors influence the uncertainty.

The variances are of course proportional to σ^2. However, note also that all terms except the last one of (13) are due to the calibration uncertainty, and that they are proportional to $c_{rr}^2 (D^{-1})_{rr}$, which only depends on the concentrations design of the calibration.

6 A simulation study

In the previous sections approximate results have been given for the precision of various combinations of predictor and estimator. These formulae can be programmed, but are complicated, and it is difficult to get an idea of how the various methods differ in precision and how the precision is influenced by various factors. Also, it is crucial to know when the formulae are reliable. The latter question has two aspects: Under what circumstances are the approximations adequate, given that the model is adequate, and how robust are the formulae to deviations from the model assumptions. We elucidated these questions, the robustness one excepted, by a designed simulation study, with six factors varied on two levels each.

6.1 Design

From the precision formula for the naïve predictor with SVD estimation, Result 5.1, we see that the precision depends on the following characteristics:

- σ, size of noise in data,
- calibration design, expressed through D^{-1}, incorporating choice of concentrations, number of specimens and quality of the calibration set,
- $\{\alpha_r, \beta_r, \gamma_{rr}\}$, the true profiles, which are influenced by the numbers of wavelengths and chromatographic times (to be called the density), and the degree of overlap between profiles,
- c_{rr}, the true concentration of constituent r.

To study the performance of various predictor–estimator combinations we simulated an LC–UV-dressed example with two constituents ($R = 2$). In accordance with the above discussion, and to simplify comparisons with the simulation study in Linder and Sundberg [3] we investigated the following six factors, at two levels each, in a 2^{6-1} factorial design:

Factor	− level	+ level
1) σ = size of noise in data	0.00125≈ 5%	0.0025≈ 10%
2) Design of the calibration	more informative	less informative.
3) Density, i.e. the size of the spectrochromatogram	20×20	10×10
4) K = number of specimens	4	2
5) degree of overlap in the true chromat. profiles	moderate	strong
6) degree of overlap in the true spectral profiles	moderate	strong

These factor levels are the same as in the simulation study in Linder and Sundberg [3], except for σ, which was reduced such that the present higher level + is the same as the previous lower level −. Note that we have not varied the number of constituents, which was fixed at $R = 2$. Nor did we take the true concentrations of the new specimen to be a factor, but held them fixed throughout the simulation study, at (c_{11},c_{22})=(2,3). Instead we used a relative measure of precision which to a large extent should remove the difference in prediction variances due to different true concentrations. The dependence on concentration is shown in Figure 2 for the other parameters fixed at levels $(+ - - - - --)$ and the two constituents at equal concentration $c_{11} = c_{22} = c$, $0 \leq c \leq 5$.

The third factor (density) determines the size of the data matrices Z, that is the product of the numbers of chromatographic times and wavelengths (20×20 or 10×10, the latter by thinning the former). For the fourth factor, K, level + ($K = 2$) is the minimum allowable number of specimens, since we had $R = 2$ constituents. The magnitude of noise in data was of relative size 5% and 10% of the average response, respectively. Figures 3 and 4 illustrate the form factors 5 and 6, the form and degree of overlap in the true, pure elution profiles and spectral profiles.

Since we had only two constituents, the choice of concentration combinations can be illustrated in the xy-plane, see Figure 5. Marks − and + represent the more and less informative calibration design sets, respectively. The point $(2, 3)$ to be predicted is also marked. The elution profiles for factor 5 were generated by a lognormal distribution for each constituent, to mimic the real behaviour with trailing peaks. Spectra for factor 6 were generated as mixtures of Gaussian peaks.

The design chosen for the study was a 2^{6-1} factorial design with the usual confounding pattern for half-fractions, that is the highest order interaction aliased with the identity. In each point of the factorial experiment, 1000 replicates were taken. As much as feasible we used the same random numbers for the white noise matrices $E^{(k)}$ over different design points.

In the simulation study we compared $6 = 3 \times 2$ different method combinations: The naïve predictor $\tilde{c}_0$, the LS predictor $\hat{c}_0$, and the refined LS predictor c_0^* were combined with each of SVD and BLLS for the calibration. To quantify the precision of prediction seen in the factorial experiment we used for each constituent a precision index defined as the natural logarithm of the relative MSE over the sample of 1000 replicates, and we compared with the corresponding theoretical relative variance formula. Relative MSE (or variance) means that it was normed by c_{rr}^2. The reason for this scaling is found in the theoretical variances, where the calibration part is proportional to c_{rr}^2. Finally, since the variance formulae are multiplicative in σ^2, taking the log is necessary to avoid large interaction effects, at least for the σ factor.

6.2 Simulation results

Results of the factorial experiment are primarily shown in Table 1 and in Figures 6, 7 and 8, but also in the subsequent three tables based on Table 1. Given in Table 1 is the precision index (log relative MSE) in each experimental point, with the number 12 added for convenience, together with the correction in log-values required to go from the theoretical approximate variance to the observed simulation MSE, denoted *Diff*. Equivalently Diff may be expressed as the amount of MSE underestimation if the variance formula is used. The results for constituents $r = 1$ and $r = 2$ were similar, and Table 1 shows only the average over them. The results were also similar for the LS and refined LS predictors, and only the latter is tabulated. More precisely, the refinement of LS gave a reduction in precision index of around 0.06 with BLLS and 0.16 with SVD. The LS and refined LS predictors were also highly mutually correlated in their variation between simulations, from almost 90% to more than 99% in many design points. Their correlation with the naïve predictor, on the other hand, was not very high. The constant 12 was only to make all the former entries positive and does not affect the *diff* columns or the effects tabulated below. When the correction value is small (within ± 0.2, say, for negligible error in the approximation $\log(1 + x) \approx x$), it can be interpreted as an estimate of the relative increase in the variance formula value required to yield the actual MSE. Half this value is then the corresponding relative increase in standard error. Analogously for other differences, so for example did the refinement of LS reduce the standard error by only about 3 % with BLLS and about 8 % with SVD.

6.2.1 Accuracy of variance formulae

The Diff columns of Table 1 show that the theoretical variance is never really higher than the observed MSE, but much lower in some points of the factorial design. It means there is a risk for serious underestimation of the size of the prediction errors under some circumstances. These design points partially differ between methods, but before going into details, let us make the reassuring observation that the difference only in few cases exceeds ≈ 0.20. The relative difference in standard errors is about half that in variances, so if the variance underestimates the MSE by Diff $= 0.20$, the standard error correspondingly underestimates by about 10%. This quantity should be regarded also in the light of the simulation uncertainty in these values (the variance $2/1000$ for a $\chi^2(1000)/1000$ corresponds to a relative standard deviation of the magnitude 5% for the randomness in a standard error estimate).

The difference Diff is seen to be higher than 0.2 essentially only when the number of specimens is small, $K = 2$. When $K = 2$ (but not when $K = 4$), high σ and bad design contributes considerably to yield high Diff-values. The Density factor and the form-factors are less influential. The worst case is seen for "Naïve with SVD", when the highest differences exceed 1. Then the actual MSE is more than a factor e higher than indicated by the variance formula.

The reasons for the substantial differences between the observed MSE and the theoretical variances are only partially found in the variance part of the MSE. Figures 6 and 7 show that there can be a considerable prediction bias if the naïve predictor is used under unfavourable conditions, in particular when $K = 2$. The refined LS predictor, on the other hand, shows a noticeable bias only when all six factors are jointly at there most disadvantegous levels (design point 32). The predictor biases should be considered in the light of the precision in the prediction. Figure 8 shows some of the same averages as in Figure 6, but now with a vertical errorbar added, which shows $\pm$ the predictor standard deviation as estimated from the simulation sample. Figure 8 yields a striking illustration of the superior efficiency of (the refined) LS over the naïve method and shows that for the latter method the bias squared is the major part of the MSE in several of the design

	Factorial design point						Naïve with SVD		Naïve with BLLS		Refined LS with SVD		Refined LS with BLLS	
No.	σ	Design	Dens.	K	Form 1	Form 2	Obs'd	Diff.	Obs'd	Diff.	Obs'd	Diff	Obs'd	Diff.
1	−	−	−	−	−	−	3.32	0.02	3.17	0.00	1.50	-0.01	1.37	-0.01
2	+	−	−	−	−	+	6.12	0.06	5.98	0.02	3.04	-0.02	3.00	-0.01
3	−	+	−	−	−	+	5.32	0.17	4.60	0.01	1.86	-0.03	1.91	-0.02
4	+	+	−	−	−	−	5.44	0.23	4.58	0.02	3.27	-0.01	2.76	-0.01
5	−	−	+	−	−	+	5.91	0.00	5.81	-0.00	3.14	0.07	3.07	0.06
6	+	−	+	−	−	−	5.91	-0.02	5.76	-0.02	4.33	0.08	4.16	0.06
7	−	+	+	−	−	−	5.06	-0.00	4.37	-0.02	3.36	0.10	2.77	0.05
8	+	+	+	−	−	+	7.94	0.18	7.21	0.01	4.79	0.11	4.54	0.06
9	−	−	−	+	−	+	5.49	0.17	5.13	0.11	2.41	-0.02	2.29	-0.04
10	+	−	−	+	−	−	5.65	0.26	5.22	0.18	3.71	0.00	3.39	-0.03
11	−	+	−	+	−	−	5.37	0.38	3.75	0.09	2.98	0.02	1.97	-0.04
12	+	+	−	+	−	+	8.93	1.30	7.13	0.68	4.54	0.31	4.03	0.16
13	−	−	+	+	−	−	5.28	0.03	4.95	0.06	3.74	0.06	3.42	0.06
14	+	−	+	+	−	+	8.17	0.24	7.87	0.23	5.32	0.11	5.17	0.07
15	−	+	+	+	−	+	7.87	0.39	6.51	0.22	4.45	0.19	3.90	0.06
16	+	+	+	+	−	−	8.20	0.58	6.56	0.29	6.07	0.32	4.81	0.09
17	−	−	−	−	+	+	5.60	-0.00	5.57	-0.01	2.47	-0.04	2.44	-0.02
18	+	−	−	−	+	−	5.66	0.01	5.59	0.00	3.29	-0.01	3.20	-0.01
19	−	+	−	−	+	−	4.70	0.07	4.23	-0.00	2.21	-0.01	1.79	-0.02
20	+	+	−	−	+	+	7.56	0.33	7.12	0.10	4.26	0.14	3.96	0.10
21	−	−	+	−	+	−	5.50	-0.02	5.44	-0.02	3.26	0.06	3.15	0.04
22	+	−	+	−	+	+	8.23	-0.02	8.20	-0.02	5.32	0.06	5.25	0.04
23	−	+	+	−	+	+	7.12	0.02	6.89	0.01	4.23	0.12	3.90	0.08
24	+	+	+	−	+	−	7.41	0.11	6.86	-0.00	4.99	0.10	4.55	0.06
25	−	−	−	+	+	−	4.82	0.06	4.62	0.03	2.64	-0.05	2.43	-0.04
26	+	−	−	+	+	+	7.70	0.37	7.57	0.32	4.76	0.12	4.66	0.14
27	−	+	−	+	+	+	7.36	0.76	6.50	0.47	3.74	0.12	3.30	0.18
28	+	+	−	+	+	−	8.15	1.16	6.75	0.69	4.78	0.19	3.96	0.16
29	−	−	+	+	+	+	7.24	0.04	7.18	0.07	4.72	0.10	4.58	0.09
30	+	−	+	+	+	−	7.61	0.17	7.42	0.18	5.50	0.14	5.25	0.11
31	−	+	+	+	+	−	7.20	0.28	6.12	0.18	4.67	0.17	3.83	0.11
32	+	+	+	+	+	+	9.81	0.56	9.33	0.66	7.16	0.78	6.49	0.65

Table 1: Simulation results averaged over both constituents. The table gives values of 12 + the natural logarithm of the simulated relative variance of prediction, denoted *Obs'd*, and the correction in log-values required to go from the corresponding theoretical approximation to the simulated variance, denoted *Diff*. For factor levels, see Section 6.1

points.

6.2.2 Factorial structure for the observed precision indices themselves

We next present and comment on a table of grand means, main effects and largest interaction effects for the precision indices corresponding to the four different prediction–estimation methods. This Table 2 is derived from Table 1. Higher order interactions are negligible. Effects are calculated as signed mean values of the responses, so they correspond to half the change in response when changing level of a factor from $-$ to $+$. Both from theoretical calculations and from statistical analyses of the interactions we have reason to believe that the standard error of the effects is of magnitude 0.01.

Factorial effect	Naïve & SVD	Refined LS & SVD	Naïve & BLLS	Refined LS & BLLS
Grand mean (+12)	6.61	3.95	6.06	3.60
Main effect 1 (σ)	0.79	0.74	0.76	0.73
Main effect 2 (Design)	0.48	0.26	(0.09)	(0.04)
Main effect 3 (Density)	0.54	0.74	0.59	0.71
Main effect 4 (K)	0.56	0.50	0.35	0.37
Main effect 5 (form 1)	0.37	0.30	0.53	0.33
Main effect 6 (form 2)	0.66	0.18	0.73	0.30
Interaction 2*4	0.21	(0.09)	(0.07)	(0.02)
Interaction 5*6	(-0.06)	0.15	(-0.02)	0.11

Table 2: Effects > 0.10 for the simulated precision index (log observed relative MSE). Standard error of effects ≈ 0.01.

The differences in grand means will be discussed in later subsections. All factors have substantial main effects, except the Design factor (No. 2) when BLLS is used. Factors 1 (σ) and 3 (Density) belong to the dominating group for all methods. For the naïve prediction method in particular, the effects of the form factors (5 and 6) are worth noting.

Some of the effects were more or less expected from the form of the variance formulae. The theoretical variances are proportional to σ^2, which implies that the theoretical effect of this factor is $\log 2 = 0.69$. This fits the simulated effects reasonably well. From Result 5.1 we can also approximately deduce what the effects of factors 3 and 4 should be, at least for the naïve & SVD combination. A reduction of the number of measurements in each profile to the half by thinning should yield approximately the same variance inflation effect as doubling σ, main effect $\log 2 = 0.69$. This turned out to fit well with the theoretical formulae, and reasonably well also with the simulated variances, as seen from Table 2. Further, halving the number K of specimens should yield the effect $0.5 \log 2 = 0.35$ for the calibration part of the variance. This is also the observed magnitude of this effect.

6.2.3 Differences Naïve–Refined LS

The differences between the naïve method and the refined LS method show a similar pattern for SVD as for BLLS estimation, so we only consider the average over the two estimation methods. The contributions to the average effects were overall (only) slightly higher for SVD than for BLLS. Table 3 gives the average difference (=grand mean), main effects of the factors, and the largest

Factorial effect type	Effect value
Grand mean	2.56
Main effect 1 (σ)	0.04
Main effect 2 (Design)	0.13
Main effect 3 (Density)	-0.15
Main effect 4 (K)	0.02
Main effect 5 (form 1)	0.13
Main effect 6 (form 2)	0.45
Interaction 2*4	0.08
Interaction 5*6	-0.17

Table 3: Main effects and interaction effects > 0.05 for the difference in precision index between naïve and refined LS prediction methods, averaged over the SVD and BLLS estimation methods and the two constituents.

interaction effects. The standard errors in these effects are not likely to be of a higher magnitude than the standard errors in the specific method effects, 0.01, from Table 2.

The large scale interpretations are quite simple. There is a quite large lack of efficiency of the naïve method. Averaged over the design points it corresponds to a variance inflation factor of $\exp(2.56) \approx 10$, and in the worst points it is more than twice as large. The form factors have the dominating influence on the variability, whereas at the other end, the amount of noise (σ) is without influence. Severe overlap of profiles as in the + setting of Form 2 makes prediction particularly more difficult for the naïve method than for an LS method. In other words, profiles must be well separated in order possibly to motivate the use of the naïve method.

6.2.4 Differences SVD–BLLS

Table 4 gives factorial effects for the difference of log MSEs between the SVD and BLLS estimation methods. The effects for the SVD–BLLS difference show similar patterns for the two prediction methods, albeit somewhat stronger for the naïve method than for the refined LS, and we restrict attention to the average over the naïve and the refined LS prediction methods. For the standard error of these effects, 0.01 is a likely magnitude.

Factorial effect type	Effect value
Grand mean	0.45
Main effect 1 (σ)	0.02
Main effect 2 (Design)	0.30
Main effect 3 (Density)	-0.01
Main effect 4 (K)	0.17
Main effect 5 (form 1)	-0.09
Main effect 6 (form 2)	-0.09
Interaction 2*4	0.10

Table 4: Main effects and interaction effects > 0.05 for the difference in precision index between SVD and BLLS estimation methods, averaged over the naïve and refined LS prediction methods and the two constituents.

The interpretation is simple here, too. Only factors 2 (Design) and 4 (K) have much influence,

and they interact. With $K = 4$ and the more informative design, little efficiency is lost by use of SVD. On the other hand, if the situation is worse in these factors, SVD is clearly less efficient, in particular if both factors are set at +. In that case SVD yields a 2.5 times higher variance. Design being the most important of the factors was also seen for profiles estimation in the previous paper on SVD and BLLS [3].

7 An illustration with fluorescence data

We illustrate the proposed methods on fluorescence data. When in fluorescence analysis a specimen is illuminated by a source of light, molecules absorb light according to their absorbance spectrum and turn into an excited state. When leaving this state they emit a specific spectrum of light, which is collected by the spectrofluorometer. The instrument is based on two monochromators. The excitation monochromator provides a range of wavelengths from the light source to the specimen. The emission monochromator collects the light emitted from the specimen. For each excitation wavelength we get an emission spectrum, proportional to the absorbance at this excitation wavelength. By varying the excitation wavelength we will register what should be a bilinear two-way structure indexed by excitation and emission wavelengths. In practice there will also be some Rayleigh light scattering, appearing as emitted light at around the excitation wavelength. This yields a part of data that is not multilinear in its structure.

We have tried the SVD and BLLS calibration methods, in combination with each of the three prediction methods, on a set of fluorescence spectra discussed by Bro [4] and available on his website. The data were generated and measured by C.A. Andersson from five specimens with different mixtures of three amino acids: tryptophan, tyrosine and phenylalanine. The original measurements used excitation wavelengths 240–300 nm and emission wavelengths 250–450 nm, both with steps 1 nm. We chose to restrict the emission wavelength interval from above to 400 nm, only to make calibration and prediction more difficult (less precise), because above 400 nm only one of the amino acids emitted. We also restricted the interval from below, to 290 nm, partly for the same reason, but thereby also avoiding most of the Rayleigh scatter. In real practice one might want to keep as much information as possible and cut off the Rayleigh scatter more selectively. We also cut off the lowest 11 wavelengths of the excitation spectra, and we utilized only each second wavelength of the spectra, so the resulting two-way data were of size 25×55 for each specimen. There seemed to be quite little white noise, but far from negligible coloured noise.

Among the five specimens we chose simply to use the first four specimens for calibration and the fifth one for prediction. The estimated excitation and emission profiles from SVD and BLLS were smooth, see Figures 9 and 10 for SVD estimates joined together as piece-wise linear curves. The estimated profiles differed quite little between the SVD and BLLS estimation methods. Also, there were hardly any detectable differences for the eye between these profiles and profile estimates constructed by PARAFAC (PLS Toolbox 2.0, from Eigenvector Research Inc.) from the first four or from all five specimens (not shown). The profiles appeared reasonable also in comparison with the spectra given by Bro [4].

In Table 5 is shown how the various predictors predicted. Firstly, there are consistently very small differences between the SVD and BLLS estimation methods. This did not come as a surprise, since these differences concern calibration, where the two methods were quite alike in their estimated profiles. For the prediction phase, the three constituents illustrate different behaviour. The naïve predictor is the least good one, between 5% and 15% above target for all three constituents. For tryptophan, the refined LS predictor is almost 10% below the target. Perhaps this should be explained as the refined LS predictor being less robust to model departures than the

other methods. For phenylalanine all predictors are seen to agree well, but they are all about 10% above target. Consistency among different predictors is not sufficient for a good prediction! The LS predictor agrees quite well with the PARAFAC predictor also for tyrosine. The latter was constructed by running PARAFAC on all five specimens jointly, followed by a proportional multiple regression calibration of the loadings in the specimen dimension on the concentration vectors of the four calibration specimens. Generally, prediction by PARAFAC should often be only little less precise that the LS and refined LS predictors, but under some circumstances it can be much worse, as demonstrated by Linder [8] under homoscedastic white noise.

We have also calculated approximate standard errors for the naïve, LS and refined LS predictors, by the precision formulae given above, and corresponding 95% prediction intervals, see Figure 11. As estimate of σ^2 we used the residual MSE around the fitted model and averaged over the four calibration specimens. The magnitude of the standard errors indicate that we should have expected smaller actual errors of prediction in most cases. There are several possible interpretations of this. The one most plausible to us is that the real error structure deviates much from the assumed one. The surface of residuals does not at all resemble white noise, but something spatially strongly correlated. The residual MSEs also differ between specimens very much more than would be expected if they represented white noise with the same σ. If σ is estimated separately from each specimen in the calibration, there is a factor 3 between the largest and smallest estimates. Another possible interpretation would be that there are systematic influences from other sources of variation. There are certainly some remainders in data from the non-bilinear Rayleigh scattering. One could also try detecting a possible fourth constituent. We did so by means of the second order advantage procedure described in Section 4, but the procedure did not point towards such an interpretation.

8 Discussion

In our approach the first phase of work is a calibration of the instrument. Our tests on real and simulated data confirm that the calibration methods BLLS and SVD yield predictors which behave quite similarly. Under homoscedastic white noise the SVD method is typically only slightly less efficient than BLLS, but clearly less efficient in badly designed calibrations, in particular if concentrations to be predicted are outside the calibration range, cf. Figure 2. Else, the lower efficiency of SVD is largely hidden behind the dominating error term from the prediction, which is the same for both predictors. A compensating advantage of the SVD method is that it does not require iterations, and under unfavourable conditions the alternating BLLS algorithm can show very slow convergence. Furthermore, predictors using SVD have more explicit variance formulae than the BLLS predictor. We therefore advocate the use of the SVD method for bilinear regression as a reasonable alternative to bilinear least squares.

The second phase is the prediction. Under the assumptions of bilinearity and homoscedastic Gaussian white noise we have demonstrated that the naïve predictor is statistically much less efficient than the LS and the refined LS predictors, and that the variance formulae well represent their uncertainty under not too bad calibration conditions, at least for the two LS-based methods. The white noise assumption is clearly an idealization, as indicated by the fluorescence example. The refined LS predictor, which most efficiently makes use of the structure of this model, is also the most likely to be vulnerable to deviations from these model assumptions, both in its construction and in its variance formulae. This was probably what the fluorescence example demonstrated.

The ordinary LS predictor is probably also sensitive in its variance formula, but less so in its construction. Now we have actually given some quite general variance formulae (the basic Result 3.3) without the white noise assumption, so if we can only model better the real noise structure,

we can also implement and calculate adequate variances. For example, we could imagine white noise replaced by coloured noise, and simple such structures might even allow simple adjustments of the formulae given. Anyhow, the general advantage of the present approach, that it makes more or less explicit variance formulae possible, contrasts with alternatives like PARAFAC.

Another type of model error is the presence of an interfering substance. One procedure to show the possibility of utilising the second order advantage to deal with this problem within the present approach has been suggested in Section 4. However, more work needs to be carried out, to try the procedure and possible competitors on more simulated and real examples, and to implement it in detail. There is also the question of how the variance formulae should be modified when such a procedure has been used to correct for an interferent.

There are several more topics that might be further studied. How do the prediction methods (incl. PARAFAC and other methods) differ in their robustness to other types of model departures? What are good diagnostics for detecting the presence of various model deviations? What are the pros and cons of the methods in these aspects, in comparison with PARAFAC and other methods based on different strategies? Finally, note again that the present approach theoretically has the advantages over PARAFAC etc. that it yields statistically somewhat more efficient predictors, that it allows the prediction uncertainty to be calculated, and that it allows the calibration and the prediction to be done at separate occasions.

Method	Name of constituent		
	Tryptophan	Tyrosine	Phenylalanine
True concentration	8.79E-07	4.40E-06	2.97E-04
Naïve SVD	5.7%	15.2%	10.1%
Naïve BLLS	5.3%	14.8%	9.8%
LS SVD	0.2%	1.1%	10.7%
LS BLLS	0.1%	1.1%	10.5%
Refined SVD	-9.4%	2.4%	11.0%
Refined BLLS	-7.6%	2.3%	10.5%
PARAFAC	-0.5%	1.0%	9.7%

Table 5: Prediction results for fluorescence data. Relative prediction errors by several different prediction methods.

Appendices

A Variance for the naïve predictor and choice of left inverse

In this appendix we prove that the Moore–Penrose inverse is optimal for minimizing the variance in the prediction, for given results from the calibration. We start with a new specimen Z_0 with unknown diagonal concentration matrix C_0, which follows the constrained bilinear model $Z_0 = AC_0(B\Gamma)^T + E_0$. From the calibration procedure we have estimates $\widehat{A}$, $\widehat{B}$ and $\widehat{\Gamma}$ of the profiles A and B and of the scaling factor Γ. To get a prediction of the unknown concentration matrix C_0 we naturally use left inverses $\widehat{A}^-$ and $(\widehat{B}\widehat{\Gamma})^-$ of $\widehat{A}$ and $(\widehat{B}\widehat{\Gamma})$, respectively, to get the equation

$$\widetilde{C}_0 = \widehat{A}^- Z_0 (\widehat{B}\widehat{\Gamma})^{-T} = \widehat{A}^- A C_0 (B\Gamma)^T (\widehat{B}\widehat{\Gamma})^{-T} + \widehat{A}^- E_0 (\widehat{B}\widehat{\Gamma})^{-T}.$$

There is an infinity of generalised inverses and we have the liberty to select one of them in some optimal way. We choose to minimize the variances of the elements of $\widetilde{C}_0$ when calibration errors are neglected, that is $\mathrm{V}[\widehat{A}^- E_0 (\widehat{B}\widehat{\Gamma})^{-T}]_{r_1 r_2}$. In the case of uncorrelated, homoscedastic noise, this variance can be calculated as

$$\begin{aligned} \mathrm{V}[\tilde{c}_{r_1 r_2}] &= \mathrm{V}\left[\sum_i \sum_j (\widehat{A}^-)_{r_1 i} e_{ij} ((\widehat{B}\widehat{\Gamma})^{-T})_{j r_2}\right] \\ &= \sum_i \sum_j \sigma^2 ((\widehat{A}^-)_{r_1 i})^2 ((\widehat{B}\widehat{\Gamma})^{-T})_{j r_2})^2 \\ &= \sigma^2 (\widehat{A}^- \widehat{A}^{-T})_{r_1 r_1} (\widehat{B}^- (\widehat{B})^{-T})_{r_2 r_2} / \gamma^2_{r_2 r_2}, \end{aligned} \tag{14}$$

We now show that expressions of the type $(M^- M^{-T})_{ij}$ for an arbitrary matrix M is minimized for $M^- = M^+$, the Moore–Penrose inverse. Any generalized inverse M^- of a full column rank matrix M can be expressed in terms of a particular generalized inverse M^* and an arbitrary matrix U, see Rao and Mitra [15], Section 2.4, page 26, and in particular we select $M^* = M^+$, which is a left inverse of M:

$$M^- = M^* + U - M^* M U M M^* = M^+ + U(I - MM^+).$$

Note that $M^+ M = I$ for the Moore–Penrose inverse of a full column rank matrix.

We will show that the diagonal elements of $M^- M^{-T}$ are minimized for $U = 0$. Utilizing that $M^+ M = I$ and that MM^+ is symmetric we find

$$\begin{aligned} M^- M^{-T} &= (M^+ + U(I - MM^+))(M^{+T} + (I - (MM^+)^T)U^T) = \\ & M^+ M^{+T} + M^+ (I - MM^+) U^T + U(I - MM^+) M^{+T} \\ & + U(I - MM^+) U^T = M^+ M^{+T} + U(I - MM^+) U^T. \end{aligned}$$

The term to be minimized is obviously $U(I - MM^+)U^T$. Since $(I - MM^+)$ is positive semidefinite so is $U(I - MM^+)U^T$ and hence its diagonal elements are non-negative, and zero for $U = 0$. Furthermore, note the simplification

$$M^+ M^{+T} = (M^T M)^{-1}. \tag{15}$$

B A prediction variance derivation

In this appendix we derive Result 3.2, which gives the prediction variances for the naïve predictor, in terms of the uncertainty in the parameter estimates from the calibration.

We start with the prediction formula (4) as expressed in the constrained model representation (2),

$$\widetilde{C}_0 = \widehat{A}^+ Z_0 (\widehat{B}\widehat{\Gamma})^{+T} = \widehat{A}^+ (AC_0\Gamma B^T + E_0)(\widehat{B}\widehat{\Gamma})^{+T}, \tag{16}$$

where $\widehat{A}$, $\widehat{B}$ and $\widehat{\Gamma}$ are estimates based on some parameter estimation method for the profiles and the scaling matrix. By error propagation methods for the Moore–Penrose inverse, see Magnus and Neudecker [16], theorem 5, page 154, we find

$$\widehat{A}^+ = (A + E_A)^+ \approx A^+ - A^+ E_A A^+ + A^+ A^{+T} E_A^T (I - AA^+),$$

and hence the simple result $\widehat{A}^+ A \approx I - A^+ E_A$. The analogous result for $(\widehat{B}\widehat{\Gamma})^{+T}$ is

$$\Gamma B^T (\widehat{B}\widehat{\Gamma})^{+T} \approx I - E_{B\Gamma}^T (\Gamma B^T)^+ \approx I - \Gamma E_B^T (\Gamma B^T)^+ - E_\Gamma \Gamma^{-1}.$$

We have then also used a product rule for the error $E_{B\Gamma}$ in $\widehat{B}\widehat{\Gamma}$, namely $E_{B\Gamma} \approx E_B\Gamma + BE_\Gamma$. Inserting these formulae in Equation (16) we obtain a first order approximation of the predicted concentration,

$$\widetilde{C}_0 \approx C_0 - A^+ E_A C_0 - C_0 \Gamma E_B^T B^{+T} \Gamma^{-1} - C_0 E_\Gamma \Gamma^{-1} + A^+ E_0 B^{+T} \Gamma^{-1},$$

which we can use for the precision calculations.

We are first interested in the variances for the diagonal elements $\tilde{c}_{rr}$ of $\widetilde{C}_0$, since these elements form the naïve predictor $\tilde{c}_0$. We get

$$\begin{aligned}
\mathrm{V}[\tilde{c}_{rr}] &\approx \mathrm{V}[-\sum_i \alpha_{ri}^+ e_{ir}^A c_{rr} - \sum_j c_{rr} e_{jr}^B \beta_{jr}^{+T} - c_{rr} e_{rr}^\Gamma/\gamma_{rr} + \sum_i \sum_j \alpha_{ri}^+ e_{ij}^0 \beta_{jr}^{+T}/\gamma_{rr}] \\
&= c_{rr}^2 \mathrm{V}[\sum_i \alpha_{ri}^+ e_{ir}^A + \sum_j e_{jr}^B \beta_{jr}^{+T} + e_{rr}^\Gamma/\gamma_{rr}] + \mathrm{V}[\sum_i \sum_j \alpha_{ri}^+ e_{ij}^0 \beta_{jr}^{+T}/\gamma_{rr}],
\end{aligned}$$

where the contribution from the new specimen (the last term) is uncorrelated with the contributions from the calibration. We require the variances for all error terms and the covariances between the first three terms. This calculation can be carried out in a straightforward way, as for example:

$$\begin{aligned}
\mathrm{V}[\sum_i \alpha_{ri}^+ e_{ir}^A] &= \sum_{i_1} \sum_{i_2} \alpha_{ri_1}^+ \alpha_{ri_2}^+ \mathrm{Cov}[e_{i_1 r}^A, e_{i_2 r}^A] = \\
&= \sum_{i_1} \sum_{i_2} \alpha_{ri_1}^+ \alpha_{ri_2}^+ \mathrm{V}[\hat{\alpha}_r]_{i_1 i_2} = (A^+ \mathrm{V}[\hat{\alpha}_r] A^{+T})_{rr}.
\end{aligned}$$

All other variances and covariances are found in the same way, with the results

$$\begin{aligned}
\mathrm{Cov}[\sum_i \alpha_{ri}^+ e_{ir}^A, \sum_j e_{jr}^B \beta_{jr}^{+T}] &= (A^+ \mathrm{Cov}[\hat{\alpha}_r, \hat{\beta}_r] B^{+T})_{rr} \\
\mathrm{Cov}[\sum_i \alpha_{ri}^+ e_{ir}^A, e_{rr}^\Gamma/\gamma_{rr}] &= (A^+ \mathrm{Cov}[\hat{\alpha}_r, \hat{\gamma}_{rr}] \Gamma^{-1})_{rr}
\end{aligned}$$

$$
\begin{aligned}
\mathrm{V}[\sum_j e^B_{jr}\beta^{+T}_{jr}] &= (B^{+}\mathrm{V}[\hat{\beta}_r]B^{+T})_{rr} \\
\mathrm{Cov}[\sum_j e^B_{jr}\beta^{+T}_{jr}, e^{\Gamma}_{rr}/\gamma_{rr}] &= (B^{+}\mathrm{Cov}[\hat{\beta}_r,\hat{\gamma}_{rr}]\Gamma^{-1})_{rr} \\
\mathrm{V}[e^{\Gamma}_{rr}/\gamma_{rr}] &= (\Gamma^{-1}\mathrm{V}[\hat{\gamma}_{rr}]\Gamma^{-1})_{rr} \\
\mathrm{V}[\sum_i\sum_j \alpha^{+}_{ri}e^0_{ij}\beta^{+T}_{jr}/\gamma_{rr}] &= \sigma^2\{(A^TA)^{-1}\}_{rr}\{(\Gamma B^TB\Gamma^{)}-1\}_{rr}\,,
\end{aligned}
$$

where for the last identity we use that the errors are homoscedastic and uncorrelated, and relation (15). Putting all of this together yields Result 3.2.

C Properties of the BLLS and SVD parameter estimation methods

We have here collected some results on the bilinear least squares (BLLS) and singular value decomposition (SVD) methods of parameter (profiles) estimation, which were briefly described in Section 5. These results are taken from Linder and Sundberg [3]. We first give results for the BLLS method, and start by the normal equations whose solution yields the BLLS estimates. As mentioned in Section 5, they may be solved by an ALS procedure, alternating between the parameter matrices A and B. The scaling matrix Γ of the constrained model (2) need not be introduced for the parameter estimation.

Result C.1: Normal equations for BLLS

The normal equations, over constituents $r = 1, \ldots, R$, can be written:

$$
\begin{aligned}
AD^{(r)}\Gamma B^T\beta_r &= T^{(r)}\beta_r \\
B\Gamma D^{(r)}A^T\alpha_r &= (T^{(r)})^T\alpha_r.
\end{aligned} \tag{17}
$$

Here the matrix statistics $T^{(r)}$ and the diagonal coefficient matrices $D^{(r)}$ are defined by

$$
T^{(r)} = \sum_{k=1}^{K} c^{(k)}_{rr}Z^{(k)},\ D^{(r)} = \sum_{k=1}^{K} c^{(k)}_{rr}C^{(k)}, \tag{18}
$$

i.e. they can be regarded as weighted sums of data matrices and concentration matrices, respectively. The matrices $T^{(r)}$ are sufficient statistics for the bilinear structure, that is they carry all information available in data about the parameters of the bilinear relationship.

We now briefly indicate how an approximate variance–covariance matrix for the BLLS estimator can be computed. Under the assumption that the estimated model falls near the true model the bilinear model may be linearized in terms of the full parameter vector $\theta = (\alpha_1, \ldots, \alpha_R, \beta_1, \ldots, \beta_R, \gamma_{11}, \ldots, \gamma_{RR})^T$ in a neighborhood of the true θ that includes the estimate $\hat{\theta}$. In order to cope with the parameter constraints, we linearize them as well. The following result, cf. Rao [17] section 4a.9, tells how to combine the linearized design matrix X with the linearized matrix H for the constraints $|\alpha_r| = |\beta_r| = 1$, to obtain the variance–covariance matrix for θ.

Result C.2: Asymptotic variance–covariance matrix of the BLLS estimator

The approximate variance–covariance matrix for BLLS is given by

$$
V[\hat{\theta}] = \sigma^2 \left\{ \begin{pmatrix} X^TX & H \\ H^T & 0 \end{pmatrix}^{-1} \right\}_{11}
$$

The interpretation of this formula is that we first take the inverse of the given matrix and then extract its upper left part, of the same size as X^TX.

Since we do not have quite explicit expressions for the variances and covariances of the BLLS estimator, we only refer to the general Result 5.2 for prediction precision formulae.

The SVD estimator is formed by singular value decomposition of suitably weighted sums of the observed data matrices and can be interpreted as obtained from reweighted least squares normal equations. The SVD estimator is constructed in the following way from the matrix statistics $T^{(r)}$ and the coefficient matrices $D^{(r)}$ defined by Equation (18). An algorithm, implemented in MATLAB, can be found at http://www.matematik.su.se /matstat /chemo /alg.html .

Result C.3: Construction of the SVD estimator
The SVD estimator is constructed from singular value decomposition of weighted data matrices by

$$\hat{\gamma}_{rr}\hat{\alpha}_r\hat{\beta}_r^T = SVD_1[\sum_{r_1=1}^{R}(D^{-1})_{rr_1}T^{(r_1)}] = usv^T,$$

which by identification of terms yields $\hat{\alpha}_r = u$, $\hat{\beta}_r = v$ and $\hat{\gamma}_{rr} = s$. Here D is the $R \times R$ matrix formed by the diagonals of the diagonal $D^{(r)}$s as rows. $SVD_1[\ldots]$ means the first singular value component of the specific constructed matrix.

We now briefly indicate how the variance–covariance matrix for the SVD estimates can be calculated. The expressions for the SVD estimator are linearized as

$$\begin{aligned}
\hat{\alpha}_r &\approx \alpha_r(1 - \alpha_r^T\xi^{(r)}\beta_r/\gamma_{rr}) + \xi^{(r)}\beta_r/\gamma_{rr} \\
\hat{\beta}_r &\approx \beta_r(1 - \alpha_r^T\xi^{(r)}\beta_r/\gamma_{rr}) + (\xi^{(r)})^T\alpha_r/\gamma_{rr} \\
\hat{\gamma}_{rr} &\approx \gamma_{rr} + \alpha_r^T\xi^{(r)}\beta_r,
\end{aligned}$$

where $\xi^{(r)} = \sum_{k=1}^{K}\sum_{r_1=1}^{R}(D^{-1})_{rr_1}c_{r_1r_1}^{(k)}E^{(k)}$. This linearization makes it possible for us to calculate variances and covariances for the estimates. In the case of uncorrelated homoscedastic noise they are particularly simple and are given by the following result.

Result C.4: Precision of the SVD estimator
For uncorrelated homoscedastic noise with variance σ^2 the approximate variances and covariances of the SVD estimator are

$$\begin{aligned}
Cov[\hat{\alpha}_{r_1}, \hat{\alpha}_{r_2}] &= \sigma^2(D^{-1})_{r_1r_2}(I - \alpha_{r_1}\alpha_{r_1}^T)\beta_{r_1}^T\beta_{r_2}(I - \alpha_{r_2}\alpha_{r_2}^T)/\gamma_{r_1r_1}\gamma_{r_2r_2} \\
Cov[\hat{\alpha}_{r_1}, \hat{\beta}_{r_2}] &= \sigma^2(D^{-1})_{r_1r_2}(I - \alpha_{r_1}\alpha_{r_1}^T)\alpha_{r_2}\beta_{r_1}^T(I - \beta_{r_2}\beta_{r_2}^T)/\gamma_{r_1r_1}\gamma_{r_2r_2} \\
Cov[\hat{\alpha}_{r_1}, \hat{\gamma}_{r_2r_2}] &= \sigma^2(D^{-1})_{r_1r_2}(I - \alpha_{r_1}\alpha_{r_1}^T)\beta_{r_1}^T\beta_{r_2}\alpha_{r_2}/\gamma_{r_1r_1} \\
Cov[\hat{\beta}_{r_1}, \hat{\beta}_{r_2}] &= \sigma^2(D^{-1})_{r_1r_2}(I - \beta_{r_1}\beta_{r_1}^T)\alpha_{r_1}^T\alpha_{r_2}(I - \beta_{r_2}\beta_{r_2}^T)/\gamma_{r_1r_1}\gamma_{r_2r_2} \\
Cov[\hat{\beta}_{r_1}, \hat{\gamma}_{r_2r_2}] &= \sigma^2(D^{-1})_{r_1r_2}(I - \beta_{r_1}\beta_{r_1}^T)\alpha_{r_1}^T\alpha_{r_2}\beta_{r_2}/\gamma_{r_1r_1} \\
Cov[\hat{\gamma}_{r_1r_1}, \hat{\gamma}_{r_2r_2}] &= \sigma^2(D^{-1})_{r_1r_2}\alpha_{r_1}^T\beta_{r_1}^T\beta_{r_2}\alpha_{r_2},
\end{aligned}$$

where I is the identity matrix of suitable size. In particular the variances and covariances within

constituent are

$$\begin{aligned} V[\hat{\alpha}_r] &= \sigma^2(D^{-1})_{rr}(I-\alpha_r\alpha_r^T)/\gamma_{rr}^2 \\ V[\hat{\beta}_r] &= \sigma^2(D^{-1})_{rr}(I-\beta_r\beta_r^T)/\gamma_{rr}^2 \\ V[\hat{\gamma}_{rr}] &= \sigma^2(D^{-1})_{rr} \\ Cov[\hat{\alpha}_r,\hat{\beta}_r] &= Cov[\hat{\alpha}_r,\hat{\gamma}_{rr}] = Cov[\hat{\beta}_r,\hat{\gamma}_{rr}] = 0. \end{aligned}$$

Remark: The matrices $\mathrm{V}[\hat{\alpha}_r]$ and $\mathrm{V}[\hat{\beta}_r]$ are singular due to the constraints $|\hat{\alpha}_r| = |\hat{\beta}_r| = 1$. For derivation of these variances and covariances, see Linder and Sundberg [3]. Note that within constituent many covariances are zero, so the correlation structure is relatively simple. Also covariances between constituents will typically be relatively small. However, all covariances will vanish only if all profiles are mutually orthogonal, a situation not likely to occur in practice. ■

Finally in this section, a derivation is given of the prediction variance for the naïve predictor with the SVD estimator, found as Result 5.1.

Justification of Result 5.1: Inserting the variance–covariance expressions from Result C.4 in Result 3.2 yields

$$\begin{aligned} \mathrm{V}[\tilde{c}_{rr}] &\approx \mathrm{V}[C_0 - A^+E_AC_0 - C_0\Gamma E_B^T(\Gamma B^T)^+ - C_0E_\Gamma\Gamma^{-1} + \widehat{A}^+E_0(\widehat{B}\widehat{\Gamma})^{+T}]_{rr} \qquad (19) \\ &= c_{rr}^2\{(A^+[\sigma^2(D^{-1})_{rr}(I-\alpha_r\alpha_r^T)/\gamma_{rr}^2]A^{+T})_{rr} + \\ &\quad (B^+[\sigma^2(D^{-1})_{rr}(I-\beta_r\beta_r^T)/\gamma_{rr}^2]B^{+T})_{rr} + \\ &\quad (\Gamma^{-1}[\sigma^2(D^{-1})_{rr}]\Gamma^{-1})_{rr}\} + \\ &\quad \sigma^2(A^+A^{+T})_{rr}(\Gamma^{-1}B^+B^{+T}\Gamma^{-1})_{rr}\,. \qquad (20) \end{aligned}$$

The expression of Result 3.2 is especially simple for the SVD estimator since all covariances within constituent are zero. This expression simplifies to

$$\mathrm{V}[\hat{c}_{rr}] = \frac{\sigma^2c_{rr}^2(D^{-1})_{rr}}{\gamma_{rr}^2}\{(A^+A^{+T})_{rr} + (B^+B^{+T})_{rr} - 1 + \frac{(A^+A^{+T})_{rr}(B^+B^{+T})_{rr}}{c_{rr}^2(D^{-1})_{rr}}\}.$$

The simplification follows from the form $M^+ = (M^TM)^{-1}M^T$ for the Moore–Penrose inverse for a full column rank matrix. It is easily checked that this formula implies $M^+M = I$ and $M^+M^{+T} = (M^TM)^{-1}$. By help of this we find

$$\begin{aligned} [M^+(I-m_rm_r^T)M^{+T}]_{rr} &= [(M^TM)^{-1} - e_re_r^T]_{rr} \\ &= [(M^TM)^{-1} - E_{rr}]_{rr} = [(M^TM)^{-1}]_{rr} - 1, \end{aligned}$$

which are used to simplify the first two terms of formula (20). ■

References

1. E. Sanchez and B.R. Kowalski, *J. Chemometrics*, **2**, 247-263 (1988)
2. E. Sanchez and B.R. Kowalski, *Anal. Chem.*, **58**, 496-499 (1986)
3. M. Linder and R. Sundberg, *Chemometrics Intell. Lab. Syst.*, **42**, 159-178 (1998)
4. R. Bro, *Chemometrics Intell. Lab. Syst.*, **38**, 149-171 (1997)
5. S. Wold, P. Geladi, K. Esbensen and J. Öhman, *J. Chemometrics*, **4**, 41-56 (1990)
6. R. Bro, *J. Chemometrics*, **10**, 47-61 (1996)
7. A.K. Smilde, *J. Chemometrics*, **11**, 367-377 (1997)
8. M. Linder, *Manuscript*, , (1999)
9. E. Sanchez and B.R. Kowalski, *J. Chemometrics*, **4**, 29-45 (1990)
10. K. Faber, A. Lorber and B.R. Kowalski, *J. Chemometrics*, **11**, 95-109 (1997)
11. A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Ellis Horwood, Chichester (1981)
12. P.M. Kroonenberg, *Three-Mode Principal Component Analysis*, DSWO Press, Leiden (1989)
13. J.B. Kruskal, *Rank, Decomposition, and Uniqueness for 3-Way and N-Way Arrays, in: Multiway Data Analysis, R. Coppi and S. Bolasco (Eds.)*, Amsterdam (1989)
14. P.J. Brown and R. Sundberg, *J. Roy. Statist. Soc.*, **49**, 46-57 (1987)
15. C.R. Rao and S.K. Mitra, *Generalized Inverse of Matrices and its Applications*, John Wiley and Sons, New York (1971)
16. J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley and Sons, Chichester (1988)
17. C.R. Rao, *Linear Statistical Inference and its Applications*, 2nd ed., John Wiley and Sons, Inc., U.S.A (1973)

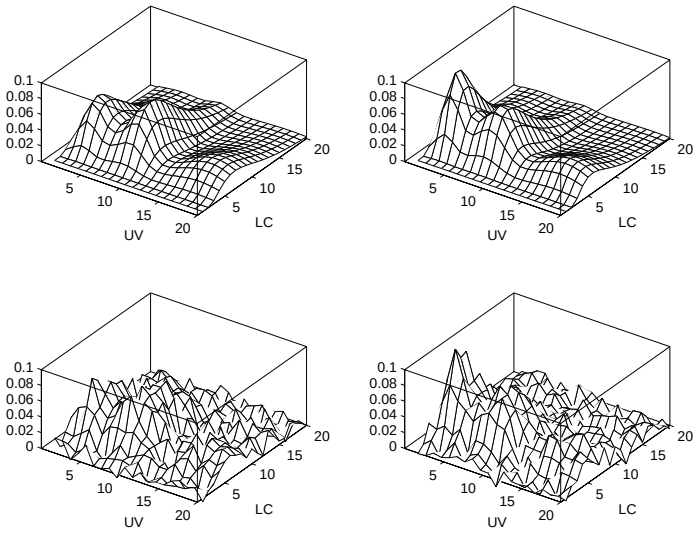

Figure 1: Pure and noisy data structures, for two specimens with the same constituents but of different concentrations.

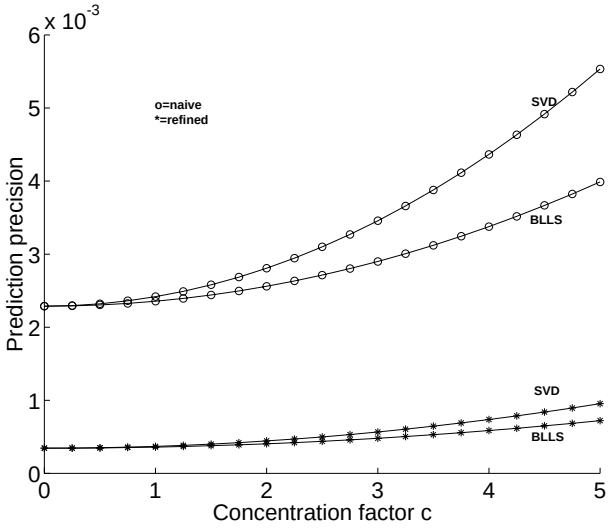

Figure 2: Example of how the precision of prediction as depends on the predicted concentration c. Theoretical variance is plotted against c for the naïve and refined LS predictors, with calibration by BLLS or by SVD, when $c_{11} = c_{22} = c$ and the factor levels are $(+ - - - - -)$.

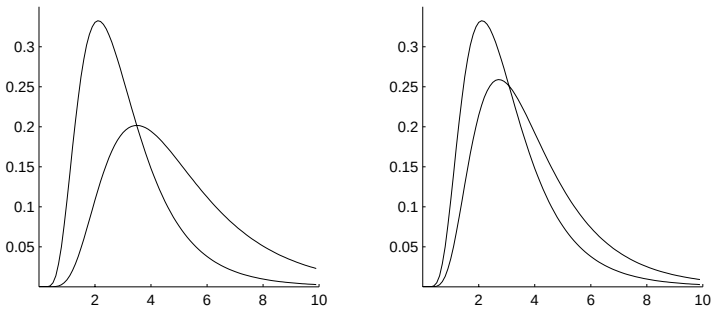

Figure 3: Factor 5 of the simulation study: elution profiles with moderate and strong overlap, respectively.

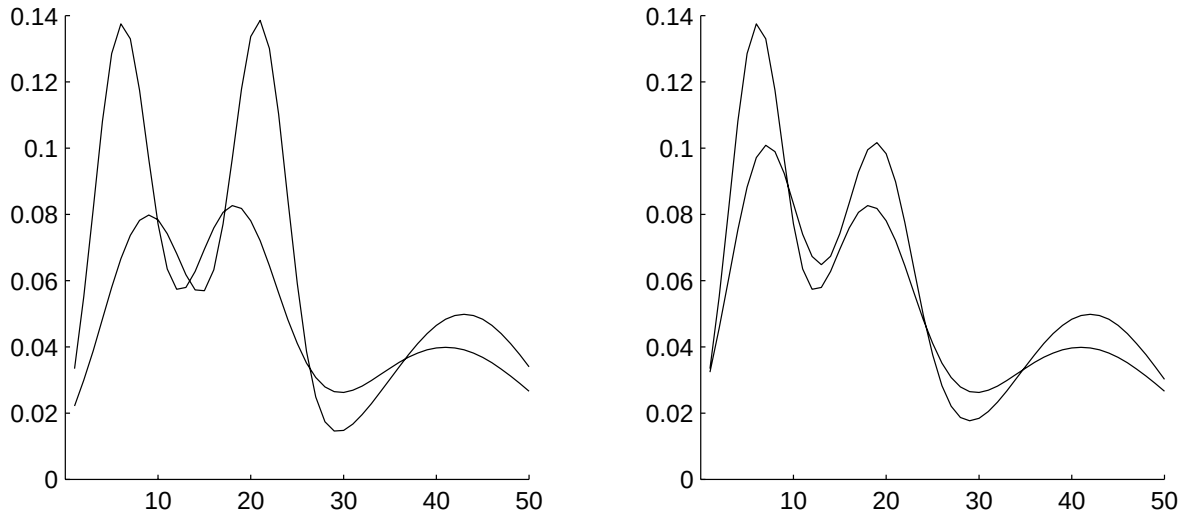

Figure 4: Factor 6 of the simulation study: spectral profiles with moderate and strong overlap, respectively.

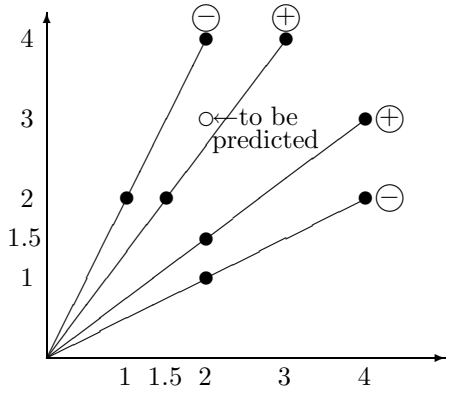

Figure 5: Calibration designs and point to be predicted in the two concentrations. Labels − on rays with points for more informative design; + on rays for less informative design. Only lower points for $K = 2$; both lower and upper points for $K = 4$

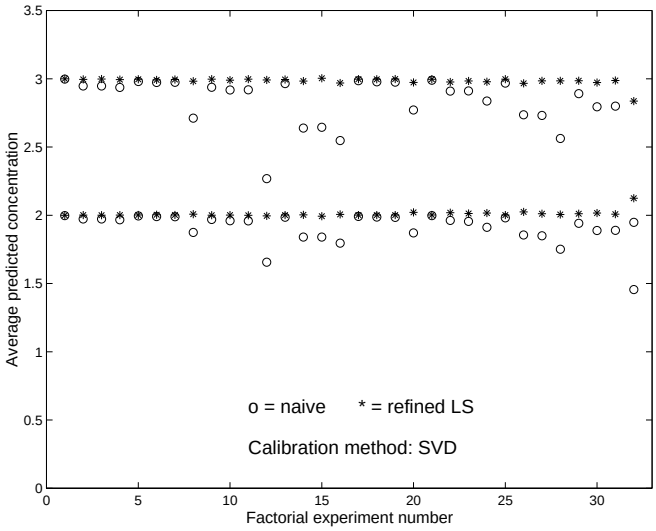

Figure 6: Average simulated naïve and refined LS predictors, with SVD for estimation, for both constituents (true concentrations 2 and 3). The x-axis represents the 32 design points of the $2^{(6-1)}$ experiment in the (standard) order of Table 1. The standard errors of the averages are typically less than half the height of the symbol (o or *)

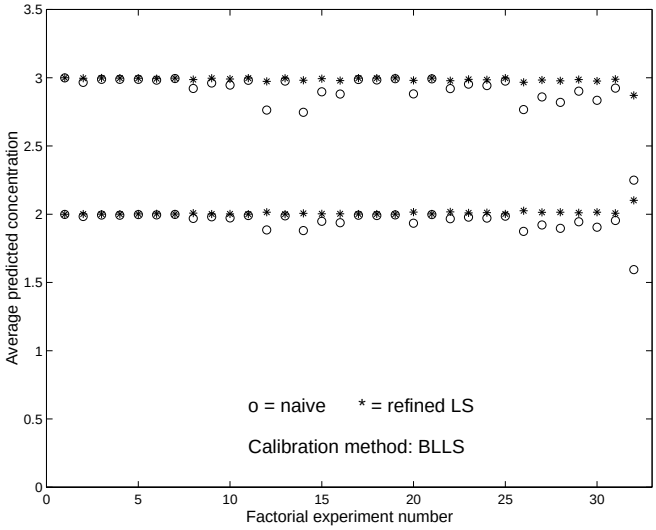

Figure 7: Average simulated naïve and refined LS predictors, with BLLS for estimation, for both constituents (true concentrations 2 and 3). The x-axis represents the 32 design points of the $2^{(6-1)}$ experiment in the (standard) order of Table 1. The standard errors of the averages are typically less than half the height of the symbol (o or *)

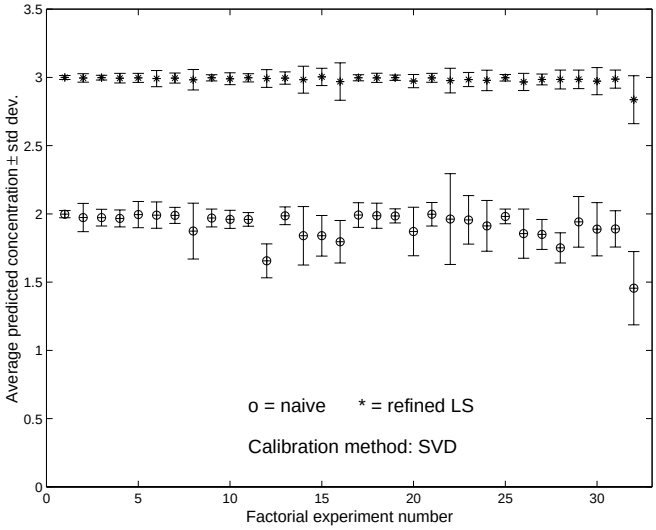

Figure 8: The naïve and refined LS predictors $\tilde{c}_0$ and c_0^* with the SVD estimator. Shown is the simulation average $\pm$ std deviation; true concentrations are 2 and 3, respectively. The x-axis represents the 32 design points of the $2^(6-1)$ experiment in the (standard) order of Table 1.

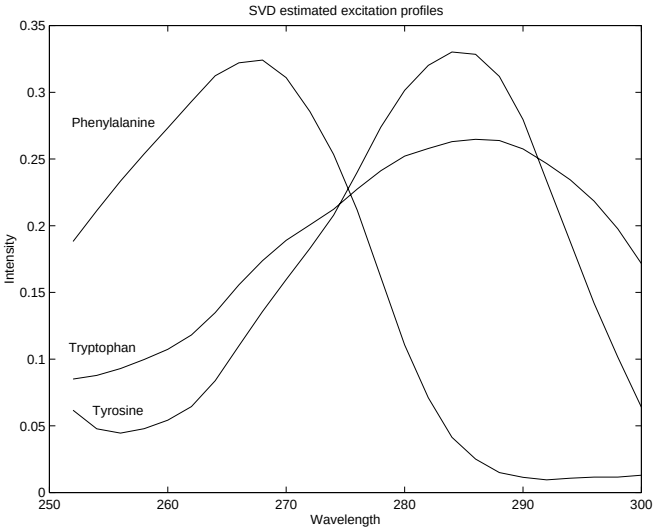

Figure 9: Amino acids exitation spectra, as estimated by the SVD method. See Section 7 for details.

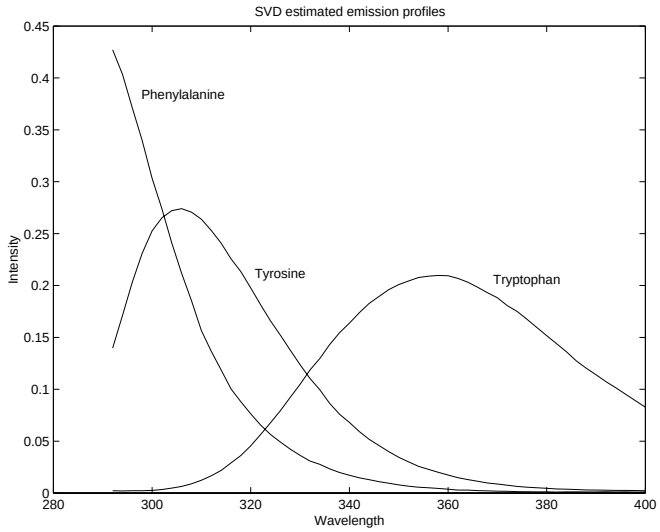

Figure 10: Amino acids emission spectra, as estimated by the SVD method. See Section 7 for details.

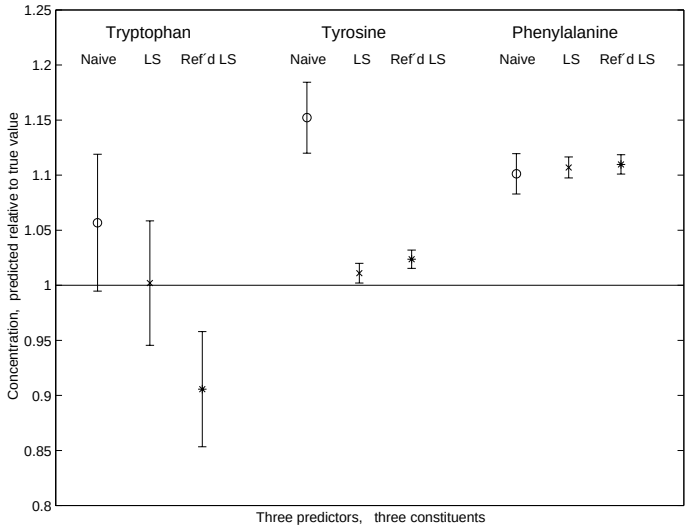

Figure 11: Prediction of three amino acids from fluorescence data. Naïve (o), LS (×) and refined LS (*) prediction methods, with errorbars representing ±2 standard errors of prediction, and scaled such that the true value corresponds to 1.